Supplementary Online Content

Ma D, Tan YJ, Ng ASL, et al. Association of *NOTCH2NLC* repeat expansions with Parkinson disease. *JAMA Neurol*. Published online August 24, 2020. doi:10.1001/jamaneurol.2020.3023

eAppendix. Supplementary Methods.
eReferences.
eTable. Details on LRS sequencing on patients with long and intermediate repeat expansions.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Supplementary Methods.

Repeat-primed PCR and repeat-size determination

Genomic DNA was extracted from peripheral blood using QIAamp® DNA Blood Maxi Kit (Qiagen, Germany). Repeat-primed ARMS PCR (RP-PCR) and fluorescence ARMS PCR were performed on all PD patients and healthy controls.

The PCR mix contained 0.75 U Taq DNA Polymerase, 1× Taq Buffer, 300 μM each dATP, dTTP, dCTP, 75 μM dGTP and 225 μM 7-Deaza-2′-deoxyguanosine-5′-triphosphate (Roche), 5% dimethyl sulfoxide (Sigma-Aldrich), 2.5 Mm MgCl2, 1 M betaine (Sigma-Aldrich), 0.4 μM each primer mix and 20 -100 ng genomic DNA in a total reaction volume of 25 μl. To determine GGC repeats, three primers were used (NOTCH2NLC-ARM-F –FAM – 5'-AGGCATTTGCGCCTGTGCTTCGAAC-3', P3- 5'-TACGCATCCCAGTTTGAGACG -3', and P4- 5'-TACGCATCCCAGTTTGAGACGGCCGCCGCCGCCGCC-3'). After initial denaturation at 96 °C for 3 min, the cycling conditions were as follows: 10 cycles of 96 °C for 30 s, 70 °C for 45 s with a reduction of 1°C per cycle, and 72 °C for 1.5 min, and 28 cycles of 96 °C for 30 s, 60 °C for 45 s and 72 °C for 1.5 min, followed by a final elongation step of 72 °C for 10 min. To determine the repeat size, two primers were used (NOTCH2NLC-ARM-F-FAM – 5'-TAGGCATTTGCGCCTGTGCTTCGAAC-3', and NOTCH2NLC-ARM-R- 5'-CAGCAGCGCCCACAGCAGAGCAGC -3') with the following PCR conditions: 96 °C for 3 min, followed by 10 cycles of 96 °C for 30 s, 70 °C for 45 s with a reduction of 1°C per cycle, and 72 °C for 1.5 min, and 25 cycles of 96 °C for 30 s, 60 °C for 45 s and 72 °C for 1.5 min, and a final elongation step of 72 °C for 10 min. Electrophoresis was performed on a 3500xl Genetic analyzer with GeneScan 500 Rox dye Size Standard (Applied Biosystems) and the data were analyzed using GeneMapper software (Applied Biosystems).

Generation of long-read sequencing (LRS) sequencing data

Genomic DNA was extracted from peripheral blood using QIAamp® DNA Blood Maxi Kit (Qiagen, Germany). LRS was performed by a commercial sequencing provider (Novogene) on the Oxford Nanopore Technology (ONT) PromethION platform using R9 flow cells. For library construction, the ONT Ligation Sequencing Kit (SQK-LSK109) was used. Input DNA were fragmented by the Megaruptor system (Diagenode) to a fragmentation length of 30 kb, and then enriched for long fragments (> 15 kb) using the BluePippin system (Sage Science). Resulting DNA were then purified, and subsequent steps such as 1) end repair, 2) dA-tailing, 3) adapter ligation and 4) motor and tether proteins ligation performed to kit manufacturer's

instructions. Upon sequencing, base-calling on raw signal data was done using Guppy 3.2.10 (ONT). NanoPack[1] was then used to remove short (< 500 bp) and low-quality (Q-score <= 7) reads.

LRS bioinformatics analysis

Using minimap2[2], we aligned raw FASTQ files from the sequencing provider to the human reference genome (hg38). Individual BAM files were then merged and sorted using Samtools[3] (version 1.9). All samples had genome-wide coverage exceeding 15X (mean = 18.9X; Supplementary Table 1), with average read lengths ranging from 4488 bp to 20811 bp (mean = 14752 bp). The Integrative Genomics Viewer[4] (IGV) was used to visualize the GGC repeat region the *NOTCH2NLC* 5' UTR (chr1:149,390,802-149,390,842) and to select reads for repeat size estimation. For samples with clearly two groups of reads based on insertion size, reads with the longer insertion size are extracted for repeat size estimation. Where all reads have approximately the same insertion size, all reads are extracted. Repeat size estimation was done by performing multiple sequence alignment using the MAFFT web service[5] in Jalview[6], and then generating a consensus sequence.

eReferences.

1. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinforma. Oxf. Engl.* **34**, 2666–2669 (2018).
2. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
3. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
4. Robinson, J. T. *et al.* Integrative Genomics Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
5. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059 (2002).
6. Clamp, M., Cuff, J., Searle, S. M. & Barton, G. J. The Jalview Java alignment editor. *Bioinforma. Oxf. Engl.* **20**, 426–427 (2004).

eTable. Details on LRS sequencing on patients with long and intermediate repeat expansions.

Patient	Depth of coverage	Average read length (bp)	# of reads used to estimate repeat size
1	21.41	14047	7
2	17.06	4488	6
3	18.14	15428	6
4	19.93	15559	7
5	19.24	18730	9
6	18.69	14696	9
7	22.40	12265	10
9	15.82	13693	5
10	18.48	20811	15
12	188.17	17804	6